

证 明

#4 3-7-01
Priority
Papers

JCS69 U.S. PTO
09/627372
07/28/00

本证明之附件是向本局提交的下列专利申请副本

申 日: 1999 07 28

申 号: 99 1 10540.0

申 别: 发明专利

发 称: 提供本国语言查询服务的方法和系统

申 人: 国际商业机器公司

发 计人: 潘越; 杨力平; 林顿·罗宾逊

Best Available Copy

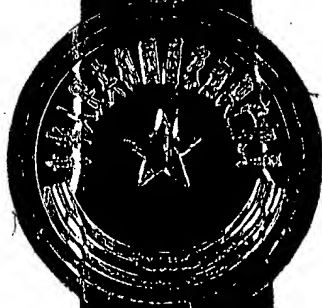
CERTIFIED COPY OF
PRIORITY DOCUMENT

中华人民共和国
国家知识产权局局长

姜 颖

2000 年 3 月 6 日

Best Available Copy



权 利 要 求 书

1. 一种为因特网用户提供本国语言查询服务的方法，该方法中利用了因特网中的多个搜索引擎，每个搜索引擎具有各自的专用语言，用于接收含有所述专用语言查询字的查询请求并且返回与所述查询字相关的查询结果，所述方法的特征在于包括以下步骤：

(a) 在一个站点接收所述因特网用户的原始查询请求，所述原始查询请求中含有所述用户的本国语言查询字；

(b) 从所述多个搜索引擎中选择一个适用的搜索引擎；

(c) 将所述本国语言查询字翻译为所选择的搜索引擎的专用语言查询字；

(d) 根据所述原始查询请求和所述专用语言查询字构造针对于所选择的搜索引擎的新查询请求；

(e) 向所选择的搜索引擎发送所述新查询请求并且接收返回的查询结果；

(f) 将所述查询结果作为与所述原始查询请求相关的查询结果返回给所述用户。

2. 根据权利要求 1 的方法，其特征在于所述步骤 (b) 包括步骤：将所述原始查询请求中的 URL 指定的搜索引擎作为所选择的搜索引擎。

3. 根据权利要求 2 的方法，其特征在于所述步骤 (c) 包括步骤：

根据所述原始查询请求中的 URL，在一个搜索引擎模板存储装置中检索一个与所述 URL 匹配的搜索引擎模板；

将所述本国语言查询字翻译为检索到的搜索引擎模板中限定的专用语言查询字。

4. 根据权利要求 3 的方法，其特征在于所述步骤 (c) 还包括步骤：

如果在所述搜索引擎模板存储装置中没有检索到与所述 URL 匹配的搜索引擎模板，则根据所述 URL，在所述站点的历史记录中，搜

索与所述 URL 对应的专用语言；

利用参数值的语言特征，判断查询字参数所在的位置；

将所述位置处的本国语言查询字翻译为所述专用语言查询字。

5. 根据权利要求 3 或 4 的方法，其特征在于所述步骤 (d) 包括步骤：将所述原始查询请求中的本国语言查询字替换为所述专用语言查询字，从而构成所述新查询请求。

6. 一种为因特网用户提供本国语言查询服务的方法，该方法中利用了因特网中的多个搜索引擎，每个搜索引擎具有各自的专用语言，用于接收含有所述专用语言查询字的查询请求并且返回与所述查询字相关的查询结果，所述方法的特征在于包括以下步骤：

(a) 在一个站点接收所述因特网用户的原始查询请求，所述原始查询请求中含有用户所请求的 URL，所述 URL 带有一个用于限定所述站点的前缀；

(b) 去除所述 URL 的前缀；

(c) 从所述多个搜索引擎中选择一个适用的搜索引擎；

(d) 如果所去除的前缀为重定向前缀，则执行以下步骤：

(d1) 向所选择的搜索引擎发送含有所述 URL 的请求并且接收作为响应的网页；

(d2) 在所述网页中，在需要输入查询字的 URL 之前增加翻译前缀，在其他 URL 之前增加重定向前缀，从而构成一个新网页；

(e) 如果所去除的前缀为翻译前缀，则执行以下步骤：

(e1) 将所述 URL 的参数中的用户本国语言查询字翻译为所选择的搜索引擎的专用语言查询字；

(e2) 用所述专用语言查询字替换所述 URL 的参数中的用户本国语言查询字；

(e3) 在所述 URL 之前增加重定向前缀；

(e4) 生成一个新网页，在所述网页中嵌入所述 URL 以及一段 Script 程序，所述 Script 程序用于使收到所述新网页的客户执行以下步骤：根据所述网页中嵌入的所述 URL 自动发出另一原

始查询请求;

(f) 将所述新网页作为与所述原始查询请求相关的查询结果返回给所述用户。

7. 根据权利要求 6 的方法, 其特征在于所述步骤(c)包括步骤: 将所述 URL 指定的搜索引擎作为所选择的搜索引擎。

8. 根据权利要求 7 的方法, 其特征在于所述步骤(e1)包括步骤: 根据所述 URL, 在一个搜索引擎模板存储装置中检索一个与所述 URL 匹配的搜索引擎模板;

将所述本国语言查询字翻译为检索到的搜索引擎模板中限定的专用语言查询字。

9. 根据权利要求 8 的方法, 其特征在于所述步骤(e1)还包括步骤:

如果在所述搜索引擎模板存储装置中没有检索到与所述 URL 匹配的搜索引擎模板, 则根据所述 URL, 在所述站点的历史记录中, 搜索与所述 URL 对应的专用语言;

利用参数值的语言特征, 判断查询字参数所在的位置;

将所述位置处的本国语言查询字翻译为所述专用语言查询字。

10. 根据权利要求 6 的方法, 其特征在于所述步骤(e)为:

(e) 如果所去除的前缀为翻译前缀, 则执行以下步骤:

(e1) 将所述 URL 的参数中的用户本国语言查询字翻译为所选择的搜索引擎的专用语言的多个查询字;

(e2) 分别用所述专用语言的多个查询字中的每一个替换所述 URL 的参数中的用户本国语言查询字, 从而形成多个 URL;

(e3) 在所述多个 URL 之前分别增加重定向前缀;

(e4) 将所述多个 URL 中的一个设定为缺省 URL;

(e5) 生成一个新网页, 在所述网页中嵌入所述多个 URL 以及一段 Script 程序, 所述 Script 程序用于使收到所述新网页的客户执行以下步骤: 以超文本形式显示所述专用语言的多个查询字, 并且根据所述网页中嵌入的所述缺省 URL 自动发出另一

原始查询请求。

11. 一种为因特网用户提供本国语言查询服务的系统，包括因特网中的多个搜索引擎，每个搜索引擎具有各自的专用语言，用于接收含有所述专用语言查询字的查询请求并且返回与所述查询字相关的查询结果，所述系统的特征在于还包括一个查询翻译服务器，所述查询翻译服务器包括：

客户接口，用于接收客户发送的查询请求，并且向客户返回查询结果；

请求分配装置，用于从所述客户接口接收所述查询请求，去除所请求的 URL 的前缀，并且根据所去除的前缀，将所述查询请求分配到不同的部件；

网页获取装置，用于从所述请求分配装置接收前缀为重定向前缀的查询请求，向 URL 指定的搜索引擎发送查询请求并且获取 URL 所请求的网页；

网页修改装置，在所获取的网页中，在需要查询字的 URL 之前加翻译前缀，在其他 URL 之前加重定向前缀，从而形成新网页，并且将所述新网页送到所述客户接口；

查询翻译装置，用于从所述请求分配装置接收前缀为翻译前缀的查询请求，将所请求的 URL 中的用户本国语言查询字翻译并替换为搜索引擎的专用语言查询字，并且在 URL 之前加重定向前缀；

网页生成装置，用于生成一个新网页，在网页中嵌入 URL 以及一段 Script 程序，该段程序使收到该网页的客户根据所嵌入的 URL 自动发出另一查询请求，并且将所述新网页送到所述客户接口。

12. 根据权利要求 11 的系统，其特征在于所述本国语言查询字为语音查询字。

13. 一种为数据库用户提供本国语言查询服务的系统，包括多个数据库，每个数据库具有各自的专用语言，用于接收含有所述专用语言查询字的查询请求并且返回与所述查询字相关的查询结果，所述系统的特征在于还包括一个查询翻译服务器，所述查询翻译服务器包

括:

客户接口, 用于接收客户发送的查询请求, 并且向客户返回查询结果;

查询翻译装置, 用于将所述客户接口收到的查询请求中的用户本国语言查询字翻译并替换为数据库的专用语言查询字;

查询结果获取装置, 用于向所述查询请求指定的数据库发送已翻译的查询请求并且获取查询结果。

14. 根据权利要求 13 的系统, 其特征在于所述本国语言查询字为语音查询字。

提供本国语言查询服务的方法和系统

本发明涉及信息检索方法和系统，尤其涉及为因特网用户或数据库用户提供查询服务的方法和系统，并且更具体地涉及为因特网用户或数据库用户提供本国语言查询服务的方法和系统

诸如远程网络这样的计算机化的信息源的发展允许数据处理系统的用户与其他服务器和网络连接，并从而检索以前在电子介质中不能获得的大量电子信息。这种电子信息正在越来越多地代替象报纸、杂志甚至是电视这样的传统信息传送媒介。

在通信中，能够彼此区别开的计算机网络通过“网关”结合在一起，网关处理从发送网络到接收网络的数据传输以及报文转换。网关是用来连接不同的网络（即，使用不同通信协议的网络）的装置，使得电子信息能够从一个网络传送到另一个网络。网关传输电子信息，把这种信息转换为与第二个网络使用的协议兼容的形式以便于传输和传递。

近年来通常使用的一种远程网络是因特网。因特网是全球最大的、开放的、由众多网络相互连接而成的计算机网络。它是由美国阿帕网（ARPAnet）发展而成的。主要采用TCP/IP协议。TCP/IP是“传输控制协议/互联网协议”的首字母缩略语，是美国国防部开发的用于计算机之间通信的软件协议。可以将因特网描述为由执行允许用户交互共享网络信息的联网协议的计算机互连而成的地理上分布的远程计算机网络系统。因为这种广泛使用的信息共享，诸如因特网这种远程网络已经非常全面地发展为“开放”系统。对于该系统，开发者能够基本上没有限制地设计软件应用程序以便执行特定的操作或者服务。对于有关因特网的节点、对象和链接的细节，可参见教科书《通

晓因特网-Mastering the Internet》，作者为 G.H.Cady 等人，该书于 1996 年由美国加利弗尼亚州 Alameda 市的 Sybex 公司出版。

在数据处理网络之间传输的电子信息通常是以超文本方式提供的。超文本是一种用复杂的非顺序关联网将文本、图象、声音和动作链接在一起的信息表示方法，它允许用户通过相关的标题进行“浏览”或“漫游”。术语“超文本”是十九世纪六十年代创造的，用于描述由计算机提供的文件，展现了与书籍、胶片和语言的线性格式不同的非线性结构。

另一方面，最近引入的术语“超媒体”几乎与“超文本”是近义词，但是侧重于诸如动画、记录的声音和视频等超文本的非文本成分。通过超媒体，将图象、声音、视频或它们的任意组合综合成信息存储和检索的综合系统。超媒体以及超文本，特别是按可由用户控制选择的交互格式，围绕着提供与人类思维平行的工作和学习环境的想法被构成。这就是说，允许用户在标题之间而不是象字母表那样顺序地从一个到下一个的运动之间建立相关性。超媒体以及超文本以这样的方式被链接，即它允许用户在搜索信息期间从一个主题跳到另一个相关的主题。超链接信息包含在超媒体和超文本文件中，这些文件允许用户只要“点击”（即，使用鼠标或其他指点装置）超链接的标题就能够获得文件或访问站点。

万维网 (WWW)，或简称为“Web”，是因特网的多媒体信息检索系统。它是因特网环境下最常用的传送数据的方法。其他方法还包括 FTP（文件传送协议）和 Gopher 等等，但未达到 Web 所具有的普通性。在 Web 环境中，客户机利用超文本传输协议（HTTP）实现对 Web 服务器的事务处理，其中 HTTP 是众所周知的应用程序协议，该协议允许用户利用称为超文本标记语言（HTML）的标准页面描述语言访问各种文件（如文本、图形、图象、声音、视频等文件）。HTML 提供基础文件格式，并且允许开发者规定与其他服务器及文件的链接。在因特网应用中，利用统一资源定位符（URL）规定至某服务器的网络路径，URL 具有定义网络连接的专用语法。

URL 典型地为这样的格式: `http://somehost/somedirectory?parameters ...`, 其中 “somehost” 是该 URL 的宿主名位置, “somedirectory” 是可以找到该 Web 网页的目录。把 URL 分解成为一个 Web 服务器的实际 IP 地址的常用方式是通过一个域名服务器。在互联网或内联网中, 域名服务器把 URL 中的宿主名变换成实际的网址。域名服务器的一个例子是目前在因特网中实现的域名服务 (DNS)。Web 用户从域名服务器请求宿主名和地址的过程有时称为解析。在 TCP/IP 中, 域名服务器把宿主名解析成一个或多个 IP 地址的列表, 这些 IP 地址回送给 HTTP 请求上的 Web 客户。每个 IP 地址规定一个服务器, 该服务器用来处理由浏览器发出的请求内容。

采用超文本协议的万维网遵循客户/服务器结构。万维网服务的客户端采用浏览器程序, 可向万维网服务器发送各种请求, 并且对服务器返回的由 HTML 语言定义的超文本信息和各种多媒体数据格式进行解释、显示和播放。

随着成千上万的公司、大学、政府机关、博物馆及市政当局发布“主页”, 因特网已经成为非常宝贵的信息资源。即使一个新用户, 只要很少的练习, 也能够浏览成百万个 Web 网页或成千个新闻组。因特网访问和相关的应用市场正在迅速扩展。

因此, 可以说万维网提供了一个“信息海洋”。在这个“信息海洋”中, 信息是以超文本网页的形式组织的。网页数量如此之多, 以致于根本不可能逐个站点、逐条链接地检索。如何在万维网中有效地、快速地和方便地检索到所感兴趣的信息, 对用户来说是十分重要的。

目前有多种信息检索工具。一种信息检索工具是目录服务。在目录服务中, 将站点/网页按类进行划分, 每类又进一步划分为更小的子类。对这些类和子类按层次结构组织起来, 并且以网页的形式显示各类或子类。在这些网页中, 含有指向子类的链接。通过一层一层地检索, 用户最终可以发现一组相关的站点/网页。随着站点/网页数量的迅猛增长, 目录的规模也变得非常巨大。例如, Yahoo! 目前已

有 25000 个类，可检索到的站点超过 50 万。

另一种信息检索工具是搜索引擎。在搜索引擎中，利用关键词对站点/网页进行索引，用户只要输入查询字（即，搜索引擎所能接受的某些关键词），即可通过搜索引擎检索到相关的站点/网页。因此，多数用户更趋向于使用搜索引擎，而不使用目录服务。另外，也可以对目录条目进行索引，从而利用搜索引擎进行检索。例如，在 Yahoo! 中，搜索结果既包括站点，也包括目录条目。

象因特网搜索引擎这样的免费或廉价的计算机应用软件使得人们容易查询到站点，从这些站点获得有关标题信息。当用户在因特网中冲浪时，目录服务和搜索引擎为用户提供了很大的帮助。因此，目录服务和搜索引擎往往被称为“入门站点”。著名的目录服务/搜索引擎包括 Yahoo!、AltaVista、Exite 等等。另外，新浪网（www.sina.com.cn）是一个非常著名的中文入门站点。然而，大多数网页都是用英文编写的，对网页分类和索引也使用英文。如果用户不懂英文，则他/她们便不能享受目录服务/搜索引擎所提供的便利。

为了解决这种问题，目前已开发了多种网页翻译软件，这些软件能够将网页从一种语言静态地或动态地翻译成用户的本国语言，例如可以将英文编写的网页翻译成中文，或者将德文编写的网页翻译成英文，等等。其中一些网页翻译软件在对网页进行翻译的同时，保留了网页中的 HTML 标记结构，从而用户可以毫无困难地阅读并点击以其本国语言标记的链接。另外，由于分类名称也可被翻译成用户的本国语言，所以用户能够利用其本国语言方便地使用目录服务。

然而，这种网页翻译软件不能解决因特网用户利用其本国语言进行检索和查询的问题。原因在于，这些网页翻译软件不能将用户查询命令中的用户本国语言查询字翻译成搜索引擎可接受的语言查询字。

为了解决因特网用户利用其本国语言进行检索和查询的问题，已经开发了几种查询翻译原型或解决方案。这些原型或方案向用户提供预定的表格，收集用户输入，然后将用户输入翻译成搜索引擎可接

受的语言，构造查询命令，发往特定的搜索引擎。但是，这些原型或方案具有如下所述的缺点。

第一，丧失了许多因特网搜索引擎所支持的组合查询和/或二次查询功能。这些组合查询和/或二次查询功能的目的是缩小检索范围。在组合查询的情况下，被匹配为检索条件的字串在逻辑上与其他检索条件比如待检索的类或出版日期组合在一起。在二次查询的情况下，第二次查询只在第一次查询返回的结果中进行。

第二，不能适当地处理多义字带来的问题。当将一种语言的某个文字翻译成另一种语言的文字时，常常有多种译文。前述查询翻译原型或方案不能适当处理这种多译文的情形。另外，当对各种语言尤其是中文中的多义字进行翻译时，这一问题就更复杂了。这里多义字是指具有多种含义的文字。例如，汉字“腔”（其拼音为 qiang1）具有四种含义，而每种含义又对应几个英文单词。第一种含义的“腔”可被翻译成英文“cavity”和“chamber”。第二种含义的“腔”可被翻译成英文“speech”。第三种含义的“腔”可被翻译成英文“tune”和“pitch”。第四种含义的“腔”可被翻译成英文“accent”和“tone”。汉语中有许多这样的多义字。目前的查询翻译软件

本发明旨在解决以上现有技术中存在的诸多问题。

因此，本发明的第一个目的在于提供一种为因特网用户提供本国语言查询服务的方法。

本发明的第二个目的在于提供一种为因特网用户提供本国语言查询服务的系统。

本发明的第三个目的在于提供一种为数据库用户提供本国语言查询服务的系统。

为实现上述第一个目的，本发明提供一种为因特网用户提供本国语言查询服务的方法，该方法中利用了因特网中的多个搜索引擎，每个搜索引擎具有各自的专用语言，用于接收含有所述专用语言查询字的查询请求并且返回与所述查询字相关的查询结果，所述方法的特

征在于包括以下步骤:

- (a) 在一个站点接收所述因特网用户的原始查询请求, 所述原始查询请求中含有所述用户的本国语言查询字;
- (b) 从所述多个搜索引擎中选择一个适用的搜索引擎;
- (c) 将所述本国语言查询字翻译为所选择的搜索引擎的专用语言查询字;
- (d) 根据所述原始查询请求和所述专用语言查询字构造针对于所选择的搜索引擎的新查询请求;
- (e) 向所选择的搜索引擎发送所述新查询请求并且接收返回的查询结果;
- (f) 将所述查询结果作为与所述原始查询请求相关的查询结果返回给所述用户。

本发明还提供一种为因特网用户提供本国语言查询服务的方法, 该方法中利用了因特网中的多个搜索引擎, 每个搜索引擎具有各自的专用语言, 用于接收含有所述专用语言查询字的查询请求并且返回与所述查询字相关的查询结果, 所述方法的特征在于包括以下步骤:

- (a) 在一个站点接收所述因特网用户的原始查询请求, 所述原始查询请求中含有所述用户所请求的 URL, 所述 URL 带有一个用于限定所述站点的前缀;
- (b) 去除所述 URL 的前缀;
- (c) 从所述多个搜索引擎中选择一个适用的搜索引擎;
- (d) 如果所去除的前缀为重定向前缀, 则执行以下步骤:
 - (d1) 向所选择的搜索引擎发送含有所述 URL 的请求并且接收作为响应的网页;
 - (d2) 在所述网页中, 在需要输入查询字的 URL 之前增加翻译前缀, 在其他 URL 之前增加重定向前缀, 从而构成一个新网页;
- (e) 如果所去除的前缀为翻译前缀, 则执行以下步骤:
 - (e1) 将所述 URL 的参数中的用户本国语言查询字翻译为所

0007-1226/96/0005-0000\$10.00/0



•

100

•

1

任

五

为实现上述第三个目的，本发明提供一种为数据库用户提供本国语言查询服务的系统，包括多个数据库，每个数据库具有各自的专用语言，用于接收含有所述专用语言查询字的查询请求并且返回与所述查询字相关的查询结果，所述系统的特征在于还包括一个查询翻译服务器，所述查询翻译服务器包括：

查询翻译装置，用于将所述客户接口收到的查询请求中的用户本国语言查询字翻译并替换为数据库的专用语言查询字；

利用本发明的方法和系统，用户只要选择了因特网中的一种搜索引擎，就能够借助于该搜索引擎所提供的内置的复杂查询功能，以其本国语言进行检索，而不必考虑该搜索引擎实际使用的语言。所以为各国的用户使用本国语言在因特网中检索和查询提供了极大的方便。

根据本发明的方法和系统，还能够进行专用于站点的翻译，并能够区别多义词的不同含义，进行多义词的翻译。

通过结合附图对本发明较佳实施方式的详细描述，本发明的上述特征和优点将会更加明显。其中

8

图 2 示出可根据本发明的最佳实施方式实现的计算机网络;

图 3 是本发明最佳实施方式中采用的客户/经纪服务器/服务器结构;

图 4 是在一个站点执行的本发明的第一种为因特网用户提供本国语言查询服务的方法的流程图;

图 5 是在一个站点执行的本发明的第二种为因特网用户提供本国语言查询服务的方法的流程图;

图 6 是本发明的为因特网用户提供本国语言查询服务的系统的示意框图;

图 7 至图 10 是利用本发明方法和系统进行中文查询时的浏览器显示举例。

图 1 是现有技术中普遍采用的客户/服务器结构。客户请求 (比如查询请求、新闻请求等) 91 由客户 92 发送到服务器 88。服务器 88 可以是在因特网或者其他通信网络上能够访问的远程计算机系统。例如,可以在用户的个人计算机 (未示出) 中运行客户 92,而在服务器 88 上设置因特网搜索引擎。

服务器 88 根据客户请求,扫描和搜索原始 (比如未压缩的) 信息 (比如在线新闻或新闻组),将过滤的电子信息作为服务器响应 93 提供给客户 92。

客户 92 可以在第一计算机中活动,服务器过程可以在第二计算机系统中活动,经过通信媒体相互通信,从而提供分布式功能并允许多个客户利用服务器的信息收集能力。

在万维网的环境下,为了检索信息,通常在客户机中运行与 HTML 兼容的浏览器。当浏览器的用户通过 URL 规定一条链路时,客户机对域名服务器发出请求,以便把 URL 中的主机名变换成服务器所处的特定网络 IP 地址。域名服务器回送一个或多个 IP 地址的清单,以应答该请求。利用这些 IP 地址中的一个,浏览器建立对某服务器的连接。若该服务器可用,其回送根据 HTML 格式化的文档或其他对象。服务器内运行相应的服务器软件,该软件把信息以 HTTP 响应的形式提供

给客户机。HTTP 响应对应于由 HTML 语言构成的 Web 页,或者其他服务器产生的数据。Web 浏览器已经成为用于访问许多网络及服务器的主要接口。

例如,当用户利用搜索引擎进行信息检索时,首先在他/她的个人计算机上运行某种浏览器软件。然后指定一个搜索引擎的 URL。比如,为了使用 Yahoo!进行搜索,用户输入 Yahoo!站点的 URL,即“http://www.yahoo.com”。Yahoo!服务器向浏览器返回主页。该主页是用上述 HTML 语言编写的,浏览器以 HTML 标记规定的方式显示主页内容。用户可以用指点装置点出所显示的主页中的超文本或输入查询字,以进行检索或查询。

这里所述的浏览器可以是市场上可获得的任何浏览器。例如,可以是网景公司的 Navigator、Communicator、微软公司的 IE,从伊利诺伊州 Urbana-Champaign 市的国家超计算应用中心(NCSA)获得的 Mosaic 牌浏览器,或是 Lynx 牌浏览器。任何其他的可提供 HTTP 所规定的功能的浏览器都可以使用。通常 Web 浏览器是一种能够对 HTML 文档进行语法分析并且显示的程序,尽管熟练的技术人员会赞同将来的浏览器能够使用诸如动态 HTML 及 XML 的其他标记语言。

典型的服务器包括一台 IBM RISC/6000 计算机(基于所谓的 RISC 工作站的精简指令集),运行 AIX (先进交互式执行程序版本 4.1 或更高)操作系统和服务器程序。通常利用拨号计算机网络,服务器从客户机接收请求,以便获得客户所请求的文件或对象,或者执行所请求的任务。在 IBM 公司的许多出版物中说明了基于 RISC 的计算机的各种型号,如《RS/6000、7013 和 7016 POWERstation 和 POWERserver 硬件技术参考手册》(订购号 SA23 - 2644 - 00)。在 IBM 公司 1985 年 11 月出版《AIX 操作系统技术参考手册》第一版以及其他出版物中说明了 AIX 操作系统。尽管以上所述是实际可用的,但是也可以采用任何其他适宜的硬件/操作系统/服务器组合。比如,采用微软公司的 Windows95/98/NT 操作系统的服务器结构等。

图 2 示出现有技术的计算机网络 80。计算机网络 80 代表因特

网，可以描述为根据这里所述的客户/服务器体系的公知计算机网络。当然，本领域内熟练的技术人员应认识到因特网不仅仅是根据本发明最佳实施方式采用的分布式计算机网络。也可以采用其他分布式计算机网络来实现计算机网络 80，比如本领域内公知的所谓“内联网”。

理论上，因特网是包括服务器 88 的一个大网络，这些服务器可由客户机，一般是个人计算机的用户，通过某些专用因特网接入提供商 84（比如 Internet America）或者在线服务提供商（比如 America On-Line、Prodigy、Compuserve 等等）来访问。每个客户机上可以运行浏览器（一个客户）以经过接入提供商访问服务器 88。每个服务器 88 操作所谓的“Web 站点”。在本发明中，Web 站点可以是一个提供目录服务的站点，也可以是一个提供搜索引擎的站点，还可以在其上执行本发明的为因特网用户提供本国语言查询服务的方法，从而与因特网中的多个搜索引擎一起构成为因特网用户提供本国语言查询服务的系统。

由于本发明涉及网络传输，故对网络及其操作原理的理解会有帮助的。在这里并不过分深入地描述本发明适用的网络的细节。例如，本发明可用于象因特网这样的全球性网络。

在图 1 所示的客户/服务器结构的基础上，本发明人提出一种基于因特网的客户/经纪服务器/服务器结构，如图 3 所示。实际上，在客户 92 与经纪服务器 200 以及在经纪服务器 200 与服务器 88 之间也遵从前述客户/服务器结构。在图 3 所示的结构中，客户首先向经纪服务器 200 发送请求 91。经纪服务器 200 对收到的客户请求 91 进行必要处理，将处理后的请求发送到服务器 88。服务器 88 收到请求后提供相应服务，向经纪服务器 200 返回响应 93。经纪服务器 200 收到响应 93 后，对响应做必要处理然后返回给客户 92。

执行本发明的方法的站点起到经纪服务器 200 的作用。客户 92 可以是现有技术中的各种浏览器，服务器 88 可以是现有技术中的各种提供搜索引擎的站点。

目前，每个搜索引擎都具有一种专用语言。一个搜索引擎只能接受用其专用语言表述的查询字。例如，Yahoo!只能接受用英语表述的查询字，所以Yahoo!的专用语言是英文。另一个例子是，新浪网（www.sina.com.cn）搜索引擎的专用语言是中文。用户的本国语言可以是与搜索引擎的专用语言相同或不同的任何语言。

图4是在一个站点执行的本发明的第一种为因特网用户提供本国语言查询服务的方法的流程图。该方法中，利用了因特网中的现有的多个搜索引擎，每个搜索引擎具有各自的专用语言，用于接收含有各自专用语言查询字的查询请求并且返回与查询字相关的查询结果。

步骤401，在一个站点接收一个因特网用户的原始查询请求，该原始查询请求中含有用户本国语言查询字。步骤402，根据原始查询请求，从多个搜索引擎中选择一个适用的搜索引擎。步骤403，将本国语言查询字翻译为所选择的搜索引擎的专用语言查询字。步骤404，根据原始查询请求和专用语言查询字构造针对于所选择的搜索引擎的新查询请求。步骤405，向所选择的搜索引擎发送新查询请求并且接收返回的查询结果。步骤406，将收到的查询结果作为与原始查询请求相关的查询结果返回给用户。

可以用Java语言编写实现图4所示查询翻译方法的软件，在多种操作系统环境下运行，包括前述AIX、Windows 95/98/NT等。由于这样的程序代码符合Servlet API标准，所以可以移嵌到许多网站中与其他Web服务器一起运行。当然，也可以在图2所示的网络中设置一个新站点，专门执行图4所示的为因特网用户提供本国语言查询服务的方法。这种情况下，可以将这个新设的站点称为查询翻译服务器。

以下说明图4中各步骤的实现方式。

图4中步骤401的实现可以有多种方式。

一种最简单的方式是，使本发明的查询翻译服务器作为一个代理服务器运行。这种方式要求用户对其客户软件的网络配置进行修

改,使得该客户软件发出的所有查询请求都发往本发明的查询翻译服务器。这种方式不但对于用户来说是不方便的,并且加重了查询翻译服务器和计算机网络的负担,因为某些不必翻译的文件(比如图象和声音文件)不得不经查询翻译服务器。

另外一种方式是采用 URL 变换。在这种方式中,每次在客户发出查询请求时,自动地在所请求的 URL 之前冠以一个前缀,该前缀例如包括本发明的查询翻译服务器在因特网中的域名和路径名。从而,客户的所有查询请求都自动地转到本发明的查询翻译服务器。

特别是,如果利用后文所述的独特的 URL 变换方法,则不需对客户软件做任何改变,从而极大地方便了用户。另一处改进在于,为了减轻查询翻译服务器和计算机网络的负担,对不同类别的 URL 冠以不同的前缀。为了说明这点,首先分析客户发送的查询请求。查询请求一般包括两种。一种查询请求中含有查询字,如果这种查询字使用的语言不同于搜索引擎的专用语言,则需要首先将用户查询请求中的这种不同于专用语言的语言(如用户的本国语言)的查询字翻译为专用语言查询字,然后才能将查询请求发往搜索引擎。另一种查询请求中不含有查询字,这类请求不需要翻译,可以由本发明的查询翻译服务器直接重定向到相应的搜索引擎。因此,对这两种查询请求需要做不同的处理。在本发明中,对两种查询请求的 URL 冠以不同的前缀,即翻译前缀和重定向前缀,从而在本发明的查询翻译服务器中受到不同的处理。

为了不对客户软件做任何变动,前述冠以前缀的过程是在查询翻译服务器中实现的。简单地说,查询翻译服务器最初获取搜索引擎的主页,在主页中所含的 URL 之前增加相应的前缀(即查询翻译服务器的 URL),然后将如此修改后的主页返回给客户。在用户随后的查询中,对搜索引擎每次返回的网页都进行类似的 URL 变换。

以上简述了图 4 中的步骤 401 的实现方法。

图 4 中的步骤 402 和 403 例如可以利用本发明所提供的多个预定的搜索引擎模板实现。这些搜索引擎模板可以以各种可供查找的方

式（如文件或数据库）存储在查询翻译服务器的存储器或磁盘中。多个预定搜索引擎模板中存储了查询翻译服务器所支持的所有因特网搜索引擎的必要信息，以便于在步骤 402 和 403 中使用。

例如，可以在模板文件中存储以下搜索引擎模板：

```
yahoo=\
    url=http://search.yahoo.com/bin/search,\
    parameterName=p,\
    language=English,\
    assemblerClass=Generic,\
    dictionaryContext=
ibm=\
    url=http://www.ibm.com:8080/db2search/,\
    parameterName=q;s,\
    language=English,\
    assemblerClass=Generic,\
    dictionaryContext=ibm
...
```

该例中，前两个模板分别是 Yahoo! 和 IBM 的搜索引擎模板。

如上所示，每个搜索引擎模板中定义了该搜索引擎使用的搜索处理程序的 URL、查询命令中需要翻译的参数表、可接受查询字的专用语言、查询字组合规则和适用于专门站点的字典上下文。查询字组合规则用于限定查询字之间的组合或连接关系，例如根据不同的搜索引擎，查询字之间可以用象“+”、“-”、“AND”、“OR”或“NOT”这样的约束符连接。专门站点的字典上下文，用于限定在专门站点内查询时为了翻译查询字所应使用的特定字典。

在此需要说明的是，对模板的利用不是唯一的方式。为了能够获得因特网搜索引擎的以上必要信息，当然可以采用任何其他方式。这些都属于本领域内普通技术人员的常识，不构成对本发明的限制。

假设在步骤 401 中接收到对于 Yahoo! 站点的原始查询请求。该

查询请求中例如包括以下 URL：

`http://search.yahoo.com/bin/search?p=游戏&y=66`

在步骤 402 中，首先在模板文件中搜索与该 URL 匹配的搜索引擎模板。于是找到以下搜索引擎模板：

```
yahoo=\
    url=http://search.yahoo.com/bin/search,\
    parameterName=p,\
    language=English,\
    assemblerClass=Generic,\
    dictionaryContext=
```

该模板是 Yahoo! 的模板。于是知道客户正在利用 Yahoo! 搜索引擎进行检索。同时知道步骤 401 中接收到的 URL 的参数表为“p=游戏”，因为模板指示只有参数“p”需要翻译。搜索引擎的专用语言为英文。搜索引擎中使用的查询字组合规则的函数名为 Generic，以函数的形式限定了 Yahoo! 搜索引擎所规定的查询字组合规则。字典上下文为空格，说明使用通用字典。

在步骤 403 中，将参数表“p=游戏”中的本国语言（此处为中文）查询字“游戏”翻译成专用语言（此处为英文）查询字，如“recreation”。

另外需要说明的是，如果模板文件中没有匹配的搜索引擎模板，则对该 URL 应用缺省规则。缺省规则首先判断该 URL 所指向的搜索引擎的专用语言（相应于模板中的 language 参数），判断的根据是嵌入该 URL 的网页（客户正是从该网页发出的请求）所采用的语言，或者在所有查询翻译服务器处理过的 URL 历史记录中寻找与该 URL 最类似的一个，并检查它所返回的网页所采用的语言。将判断结果标识为专用语言。然后，判断查询字参数所在的位置（相应于模板中的 parameterName 参数），判断的根据是查找参数表中用本国语言表述的参数值。判断出查询字参数所在的位置之后，将本国语言查询字参数翻译为专用语言查询字。与模板中的 assemblerClass 参数和

dictionaryContext 参数相应的参数则取各自的缺省值。

步骤 404 根据原始查询请求和专用语言查询字构造针对于所选择的搜索引擎的新查询请求。本例中，新的查询请求中包括以下 URL：

`http://search.yahoo.com/bin/search?p=recreation&y=66`

步骤 405 可以直接在查询翻译服务器中执行，但这样将损失本发明的部分特性（例如，向用户提供多个译文选项）。另外，一种更好的方式是，事先编写某种 Script 程序（该程序将在随响应网页到达客户后由客户自动执行），利用步骤 404 中得到的查询请求以及 Script 程序组成网页，将该网页作为响应返回给客户。由于 Script 程序的特点，客户收到响应后，自动执行 Script 程序。

由于中文“游戏”可以被翻译成多个英文单词，例如“recreation”、“game”或者“play”，所以步骤 403 和步骤 404 可以产生多个 URL，例如：

`http://search.yahoo.com/bin/search?p=recreation&y=66`，

`http://search.yahoo.com/bin/search?p=game&y=66`，以及

`http://search.yahoo.com/bin/search?p=play&y=66`。

将这三个 URL 中的某一个，比如

`http://search.yahoo.com/bin/search?p=recreation&y=66`，

标识为缺省的 URL。

借助于所编写的 Script 程序，将翻译时产生的多种译文选项以适当的方式向用户显示，同时将其中缺省的已翻译的查询请求再经由查询翻译服务器发往搜索引擎。由服务器将一段 Script 程序加到网页中返回给客户，客户在收到该网页中自动执行网页中所含的 Script 程序，是 Script 语言本身提供的功能，是本领域内普通技术人员的常识，在此不必多述。

相应于步骤 401 的多种实现方式，步骤 406 也可以有相应的多种实现方式。如果步骤 401 是采用代理服务器的方式实现的，则步骤 406 只简单地将搜索引擎返回的网页返回客户即可。如果步骤 401 是

采用 URL 变换的方法实现的,则步骤 406 在对搜索引擎返回的网页进行了前述 URL 变换之后才将网页返回客户。

图 5 是在一个站点执行的本发明的第二种为因特网用户提供本国语言查询服务的方法的流程图。该方法中,利用了因特网中的现有的多个搜索引擎,每个搜索引擎具有各自的专用语言,用于接收含有各自专用语言查询字的查询请求并且返回与查询字相关的查询结果。

步骤 501 接收用户的原始查询请求。假设图 5 所示的查询翻译方法运行在某个入门站点上,该入门站点的 URL 例如为 `http://portal-site/`,则在步骤 501 收到的客户请求中,所请求的 URL 中必然含有 `http://portal-site/`。

步骤 502 去除所请求的 URL 中的前缀。如在下文所举例说明的,这里的前缀可能是

`http://portal-site/redirect/` (重定向前缀),或者

`http://portal-site/translate/` (翻译前缀)。

当然,这两种前缀不构成对本发明的限制。在实际应用中,根据不同的功能设计,可以有多种前缀,只要这些前缀符合 URL 的专用语法。

步骤 503 从多个搜索引擎中选择一个适用的搜索引擎。

步骤 504 判断所去除的前缀是否为重定向前缀, `http://portal-site/redirect/`。如果步骤 504 的判断结果为否,则处理进行到步骤 505;否则处理进行到步骤 509。由于在该例中,只包括两种前缀,即重定向前缀和翻译前缀。所以,在进行到步骤 505 之前所去除的前缀必然是翻译前缀,即 `http://portal-site/translate/`。

假如在步骤 501 收到的请求中,所请求的 URL 是

`http://portal-site/translate/http://search.yahoo.com/bin/search?p=游戏&y=66`

则在到达步骤 505 时,去除前缀后的 URL 成为

`http://search.yahoo.com/bin/search?p=游戏&y=66`

在步骤 505 中，以与图 4 中的步骤 402、403 类似的方法（例如，利用搜索引擎模板），将上述 URL 上的本国语言（该例中为中文）查询字“游戏”翻译为专用语言（该例中英文）查询字。然后，在步骤 506 中，用专用语言查询字替换 URL 的参数中的本国语言查询字。

由于中文“游戏”可以被翻译成多个英文单词，例如“recreation”、“game”或者“play”，所以步骤 506 中可以产生多个 URL，例如：

`http://search.yahoo.com/bin/search?p=recreation&y=66`，

`http://search.yahoo.com/bin/search?p=game&y=66`，以及

`http://search.yahoo.com/bin/search?p=play&y=66`。

将这三个 URL 中的某一个，比如

`http://search.yahoo.com/bin/search?p=recreation&y=66`，

标识为缺省的 URL。

为了提供对组合查询命令的支持，在步骤 505 和 506 中，可以首先根据搜索引擎模板中的查询字组合规则（`assemblerClass`）确定查询命令中的约束符，比如“+”、“-”、“AND”、“OR”或“NOT”等等，然后将约束符所分隔的本国语言查询字逐一翻译为替换为专用语言查询字。

为了提供对专门站点搜索的支持，可以首先根据搜索引擎模板中的字典上下文（`dictionaryContext`）确定一个适用的字典，然后根据这个字典对查询字进行翻译。举例来说，如果用户在 IBM 站点中查询，则可将查询字“服务器”直接翻译并替换为“RS/6000”、“AS/400”或者“S/390”。如果在 HP 站点中查询，则可将查询字“服务器”直接翻译并替换为“HP3000”或者“NetServer”等。

在步骤 506 之后，处理进行到步骤 507。在步骤 507，在所述 URL 之前增加重定向前缀。于是，得到以下三个 URL：

`http://portal-site/redirect/http://search.yahoo.com/bin/search?p=recreation&y=66`，

http://portal-site/redirect/http://search.yahoo.com/bin/search?p=game&y=66, 以及

http://portal-site/redirect/http://search.yahoo.com/bin/search?p=play&y=66。

然后, 处理进行到步骤 508。在步骤 508 中, 生成一个新网页, 在该网页中嵌入上述三个 URL 以及一段 Script 程序。该程序将由接收到该网页的客户自动执行。通过这段 Script 程序的执行, 以适当方式向用户显示中文查询字“游戏”的多个译文选项。译文选项的显示方式可以有多种。一种较好的方式是, 动态生成一个网页并在一个新开的窗口中显示。在所显示的网页中以醒目方式显示多个译文选项。每个译文选项对应一条超链接, 即含有该译文的 URL。在本例中, 中文查询字“游戏”有三个译文选项, 即英文译文“recreation”、“game”和“play”。于是, 可以在新开的窗口中显示这三个译文选项。用户可以用指点装置点击这三个译文选项中的任一个, 然后浏览器即发出相应的另一原始查询请求。前述 Script 程序的另外一个功能是, 使客户根据缺省的 URL (该例中为 http://portal-site/redirect/http://search.yahoo.com/bin/search?p=recreation&y=66) 发出新的查询请求。

另外, 在步骤 508 嵌入到网页中的 Script 程序还可以向用户提供修改搜索引擎模板的功能。例如, 向用户显示查询翻译服务器所使用的搜索引擎模板的内容, 如前文所述的四项内容。用户可以对模板内容进行修改, 并重新提交查询请求。

步骤 508 之后, 处理进行到步骤 511。在步骤 511, 将生成的新网页作为响应返回给客户。

如前文所述, 如果步骤 504 的判断结果为“是”, 则处理进行到步骤 509。在步骤 509, 经过因特网获取 URL 所标识的网页。

举例来说, 如果步骤 511 将含有 http://portal-site/redirect/http://search.yahoo.com/bin/search?p=recreation&y=66 的网页返回给客户, 则客户自动执行前文所述的 Script 程序, 发出新的

HTTP 请求。在经过步骤 501、502、503 和 504 处理之后，在步骤 509，获取以下 URL 所标识的网页：

`http://search.yahoo.com/bin/search?p=recreation&y=66`

具体的过程是这样的：将该请求发往 Yahoo! 搜索引擎，Yahoo! 搜索引擎完成搜索之后以网页形式返回搜索结果。然后，处理进行到步骤 510。

在步骤 510，对从搜索引擎收到的网页进行分析并变换其中的 URL。

作为简短背景，简要介绍一下编写网页时所使用的超文本标记语言（HTML）。HTML 语言使用 <> 符号表示的标记和/或标记组，来控制 Web 页上的显示信息。因此，所有 HTML 文件必须以 <HTML> 标记开始，并且以 </HTML> 标记结束。其他的标记例如包括：表示链接锚点和超文本引用的 <A HREF> 和 ，规定各帧内容的 <FRAME> 和 </FRAME>，将图象装入某帧的 <FRAME SRC>，以及与表格有关的 <FORM ACTION> 等。许多标记设计成能够互相嵌套，以提供增强型对象。因此，例如，通过在 <A HREF> 标记内放置 IMG SRC 标记，就能够创建可搭接的图象。当然，以上标记只是举例，因为 HTML 是一种正在进化的语言。

在前述步骤 510 中，只对网页中的三种 URL 进行变换。这三种 URL 是：

- (1) <A> 或 <AREA> 标记中的“HREF”参数；
- (2) <FRAME> 标记中的“SRC”参数；
- (3) <FORM> 标记中的“ACTION”参数。

在第 (1) 和第 (2) 种 URL 之前，增加重定向前缀：

`http://portal-site/redirect/`

在第 (3) 种 URL 之前，增加翻译前缀：

`http://portal-site/translate/`

在对网页中的 URL 进行上述变换之后，形成新网页。然后，处理进行到步骤 511。在步骤 511，将生成的新网页作为响应返回给客

户。

图 6 是本发明的为因特网用户提供本国语言查询服务的系统的示意图，图中虚框为本发明的查询翻译服务器，它与因特网 600 中的现有多个搜索引擎造成了本发明的为因特网用户提供本国语言查询服务的系统。如图 6 所示，本发明的查询翻译服务器包括客户接口 601，用于接收客户发送的查询请求并且向客户返回查询结果；请求分配装置 602，用于按照查询请求中 URL 前缀的种类，分别将 URL 前缀为重定向前缀的查询请求分配给网页获取装置 603，将 URL 前缀为翻译前缀的查询请求分配给查询翻译装置 605；网页获取装置 603 向 URL 指定的搜索引擎发送查询请求获取 URL 所请求的网页；网页修改装置 604，对网页获取装置 603 获取的网页中的某些 URL 进行变换，即在需要查询字的 URL 之前加翻译前缀，在其他 URL 之前加重定向前缀，形成新网页；查询翻译装置 605，用于将所请求的 URL 中的本国语言查询字翻译并替换为搜索引擎专用语言查询字，并且在 URL 之前加重定向前缀；网页生成装置，用于生成一个新网页，在网页中嵌入 URL 以及一段 Script 程序，该程序使收到该网页的客户根据所嵌入的 URL 自动发出另一查询请求。另外，查询翻译装置 605 在工作中还利用了字典 607 和搜索引擎模板存储装置 608。字典 607 中存储了大量字典条目，每个字典条目以相互对应的方式存储了原文文字及其译文。搜索引擎模板 608 如前文参照图 4 和图 5 所述。网页生成装置 606 在工作中利用了一个 Script 程序生成器 609，用于生成前述供客户执行的 Script 程序。

图 7 至图 10 是利用本发明方法和系统进行中文查询时的浏览器显示举例。

图 7 以举例的方式说明可以利用一个入门站点进入 Yahoo! 搜索引擎。当然，也可以不通过这样的入门站点，而采用其他的方法在 Yahoo! 的 URL 之前加指向本发明的查询翻译服务器的前缀。

图 8 显示在 Yahoo! 搜索引擎的主页中键入中文查询字“游戏”。

图 9 显示用户在图 8 中键入中文查询字“游戏”之后，借助于

本发明的查询翻译器，Yahoo!返回的网页，即查询结果。另外，浏览器借助于Script程序打开另外一个窗口，显示“游戏”的三个英文译文选项。此外，还显示“Modify Current Setting”。用户可以点击这段超文本，对Yahoo!搜索引擎的模板进行修改。

图10是在图9中点击超文本“Modify Current Setting”之后，浏览器的显示。图10中显示了查询翻译服务器当前所用的Yahoo!模板的内容。用户可以进行适当修改。

可以将本发明的方法与网页翻译服务结合起来，提供完整的检索翻译服务。这种应用的举例示于图8和图9。即，用户键入中文查询字，然后得到中文查询结果。另外，本发明还可以与语音识别系统结合起来，提供对语音查询字请求中的语音查询字进行翻译。

另外需要说明的是，尽管以上都是针对因特网搜索引擎对本发明的方法和系统做了说明，但是本发明的构思还可以应用于数据库的检索和查询。根据这种构思，可以构造一种为数据库用户提供本国语言查询服务的系统，包括多个数据库，每个数据库具有各自的专用语言，用于接收含有所述专用语言查询字的查询请求并且返回与所述查询字相关的查询结果，所述系统的特征在于还包括一个查询翻译服务器，所述查询翻译服务器包括：客户接口，用于接收客户发送的查询请求，并且向客户返回查询结果；查询翻译装置，用于将所述客户接口收到的查询请求中的用户本国语言查询字翻译并替换为数据库的专用语言查询字；查询结果获取装置，用于向所述查询请求指定的数据库发送已翻译的查询请求并且获取查询结果。类似地，在数据库查询命令中的本国语言查询字也可以为语音查询字。

虽然以上结合附图详细描述了本发明的最佳实施方式，但是对于本领域内熟练的技术人员而言，可以做出各种修改和变更，而不背离本发明的范围和实质。因此，本发明的范围仅由权利要求书限定。

说明书附图

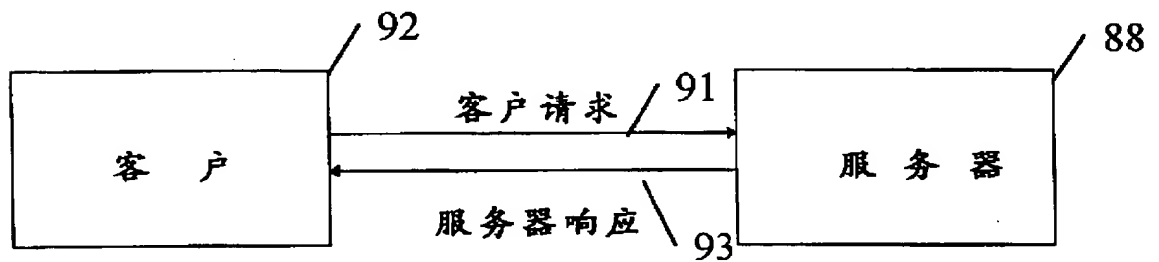


图 1

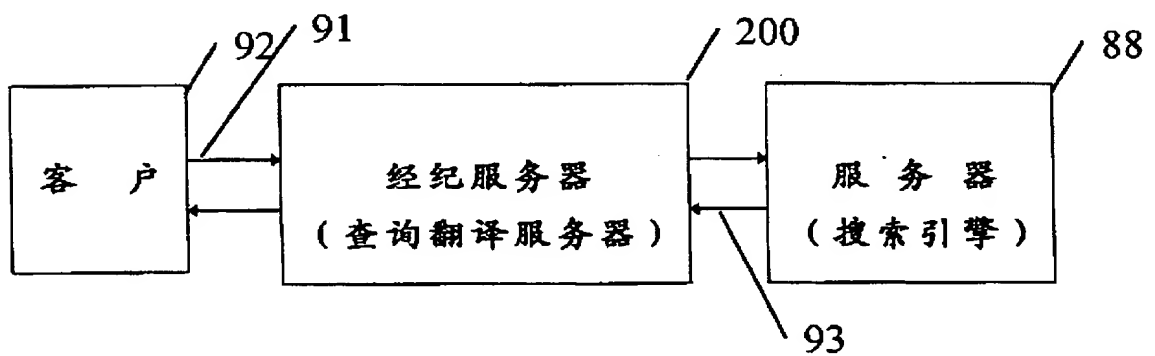


图 3

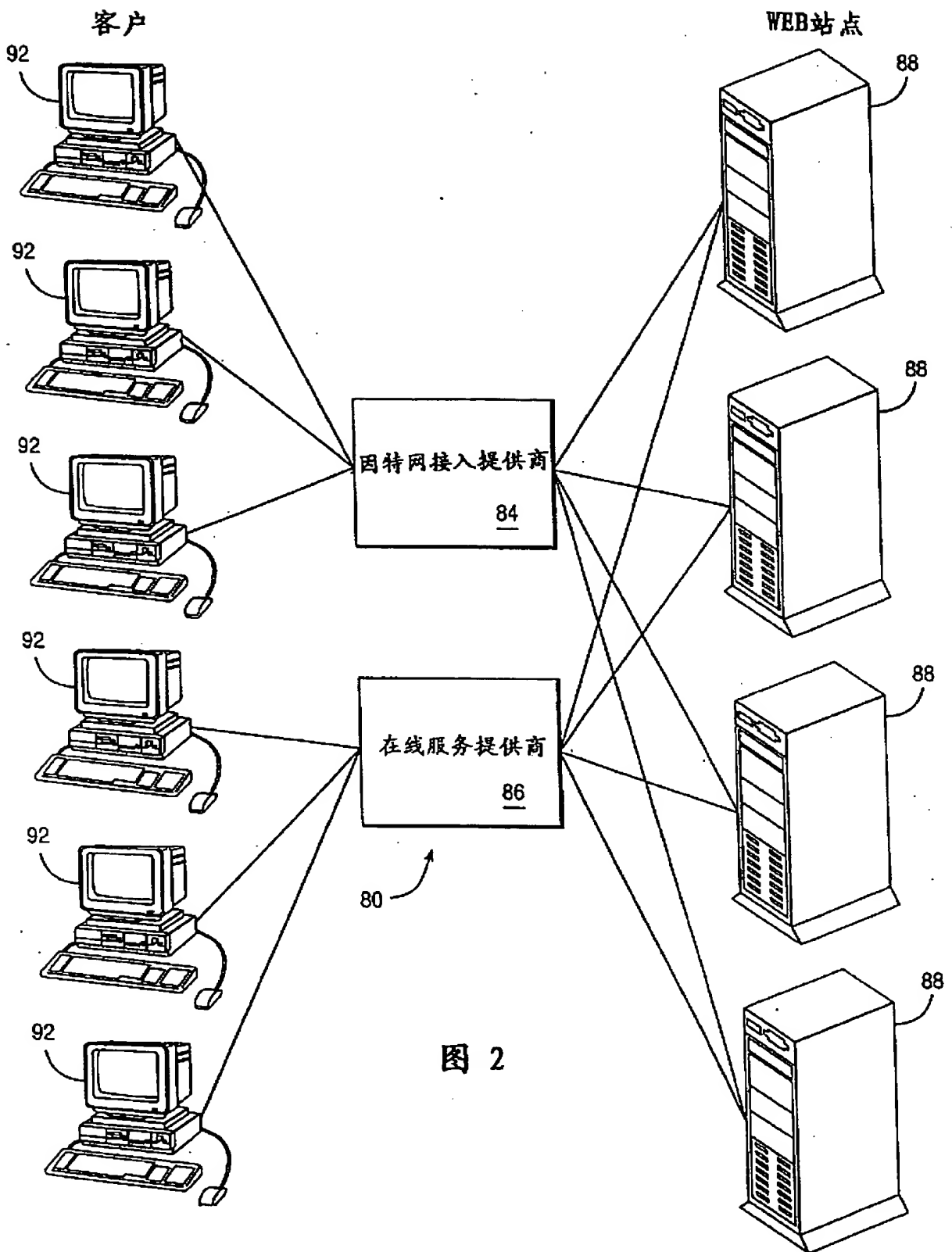


图 2

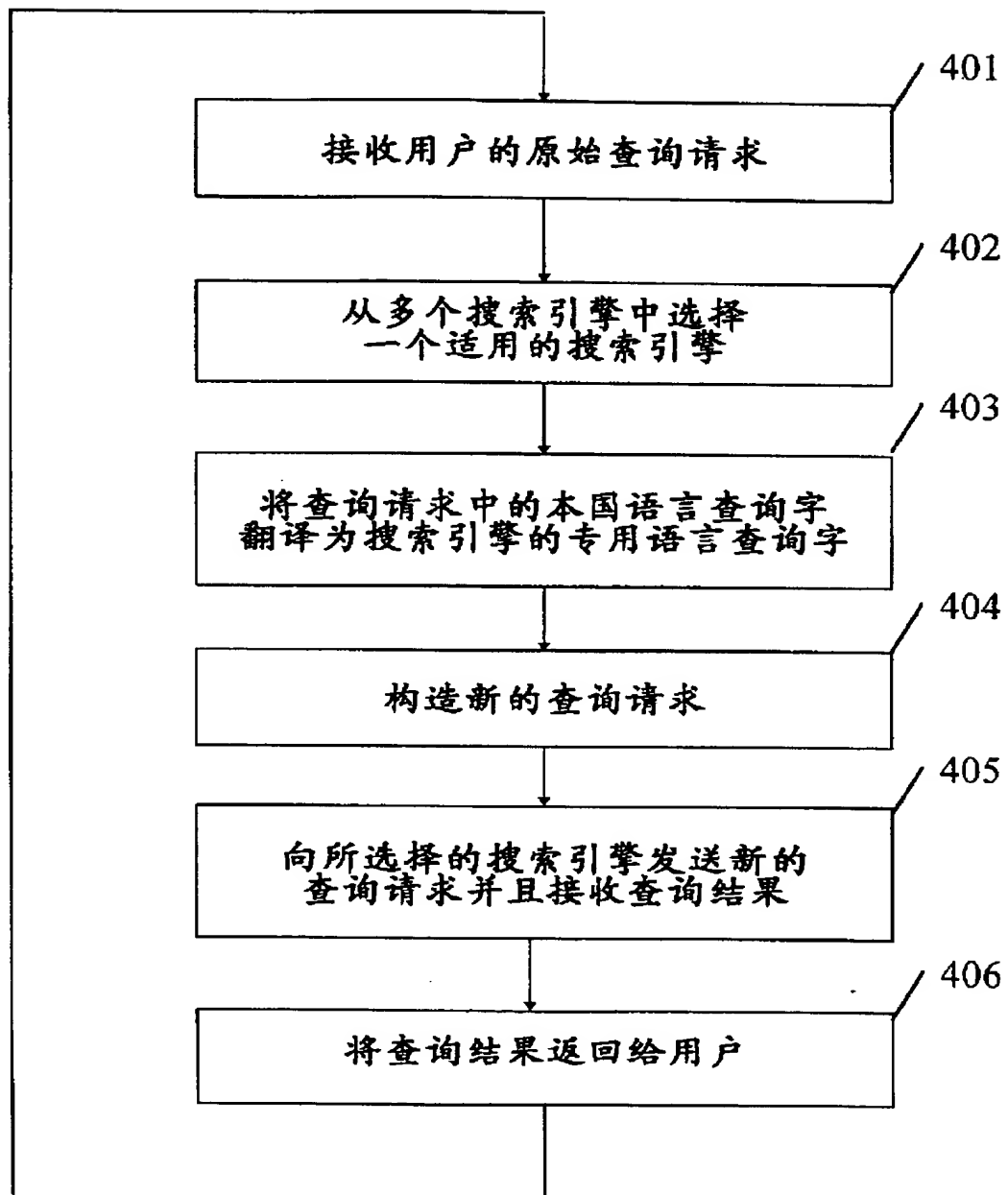


图 4

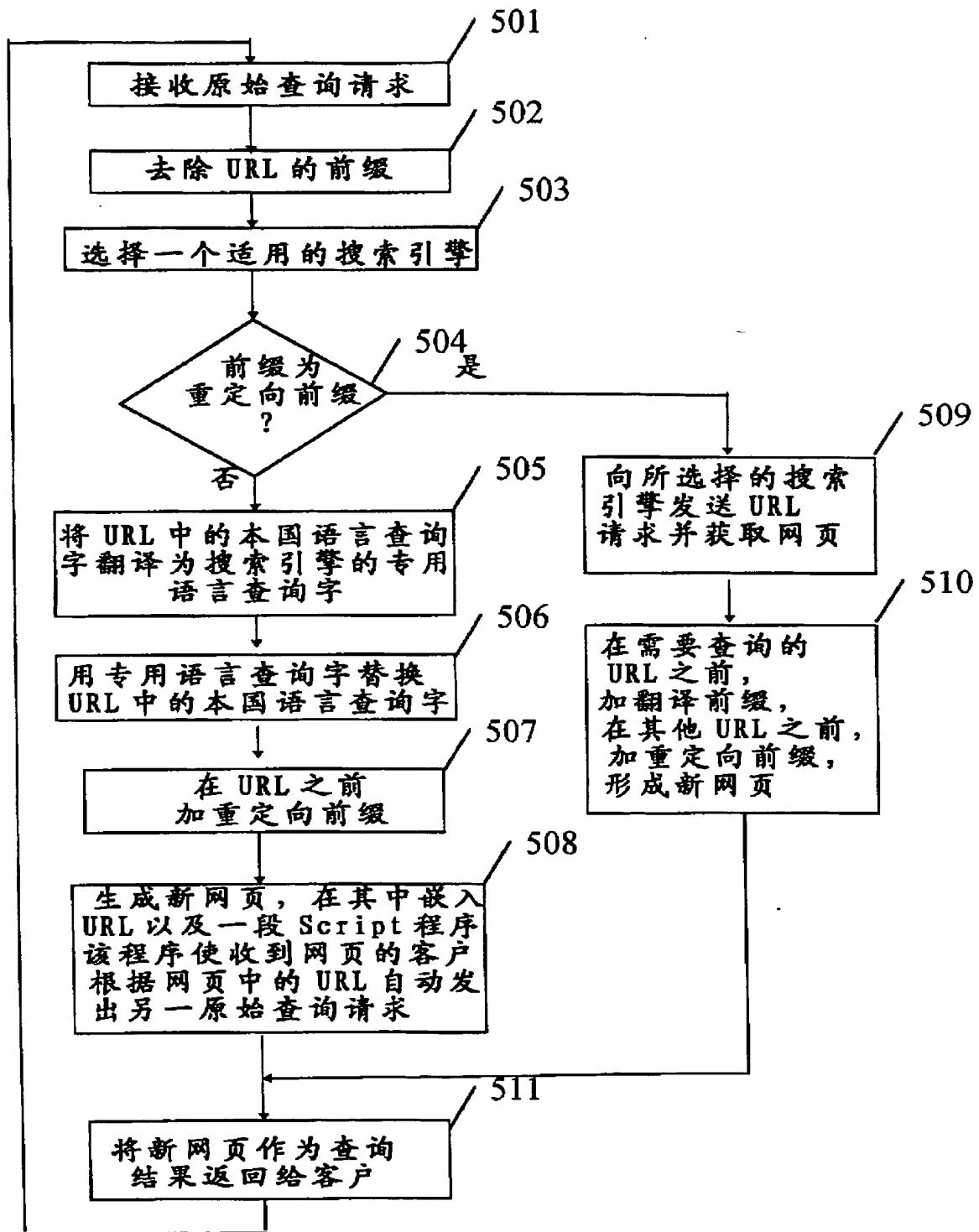
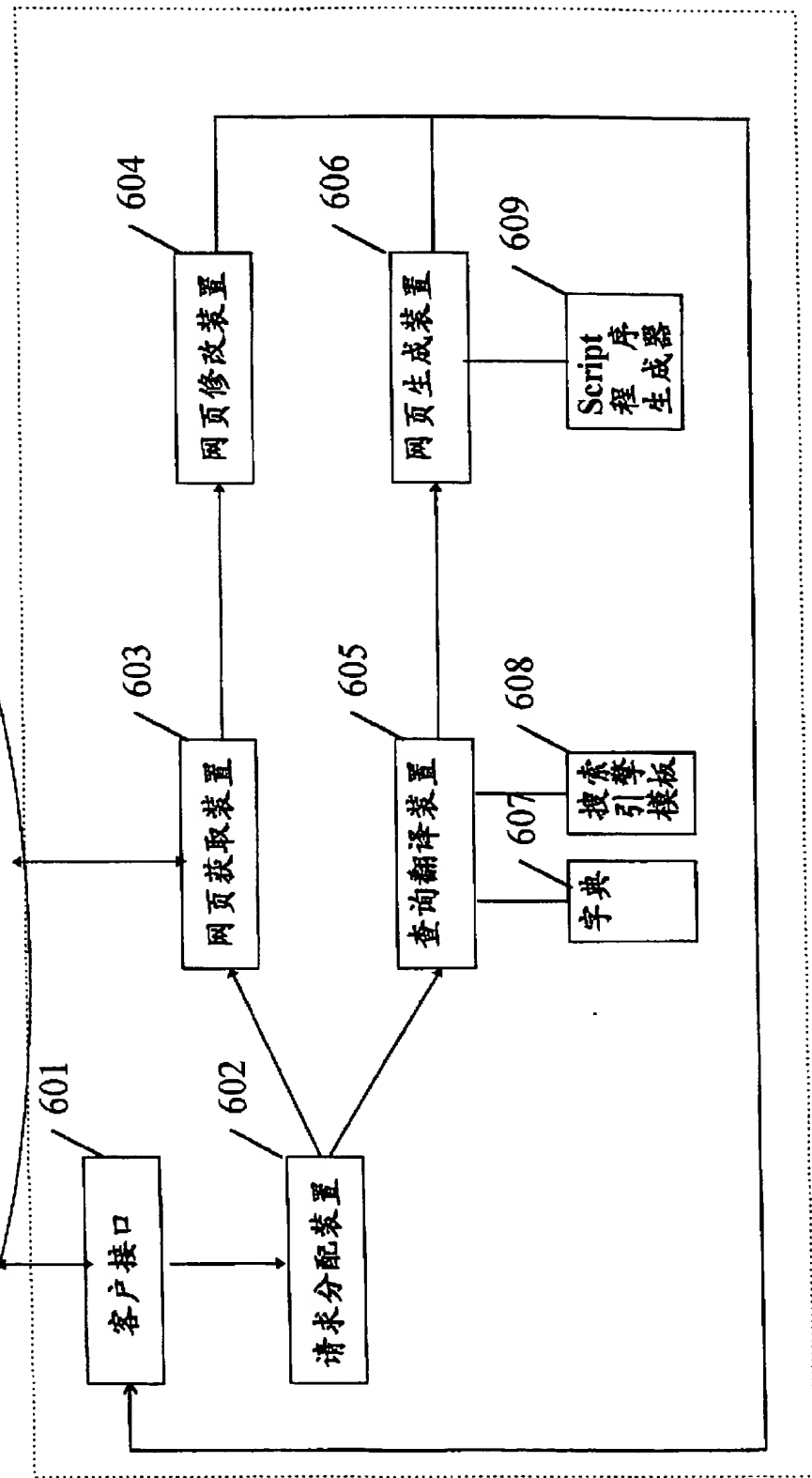


图 5

图 6
网 特 因



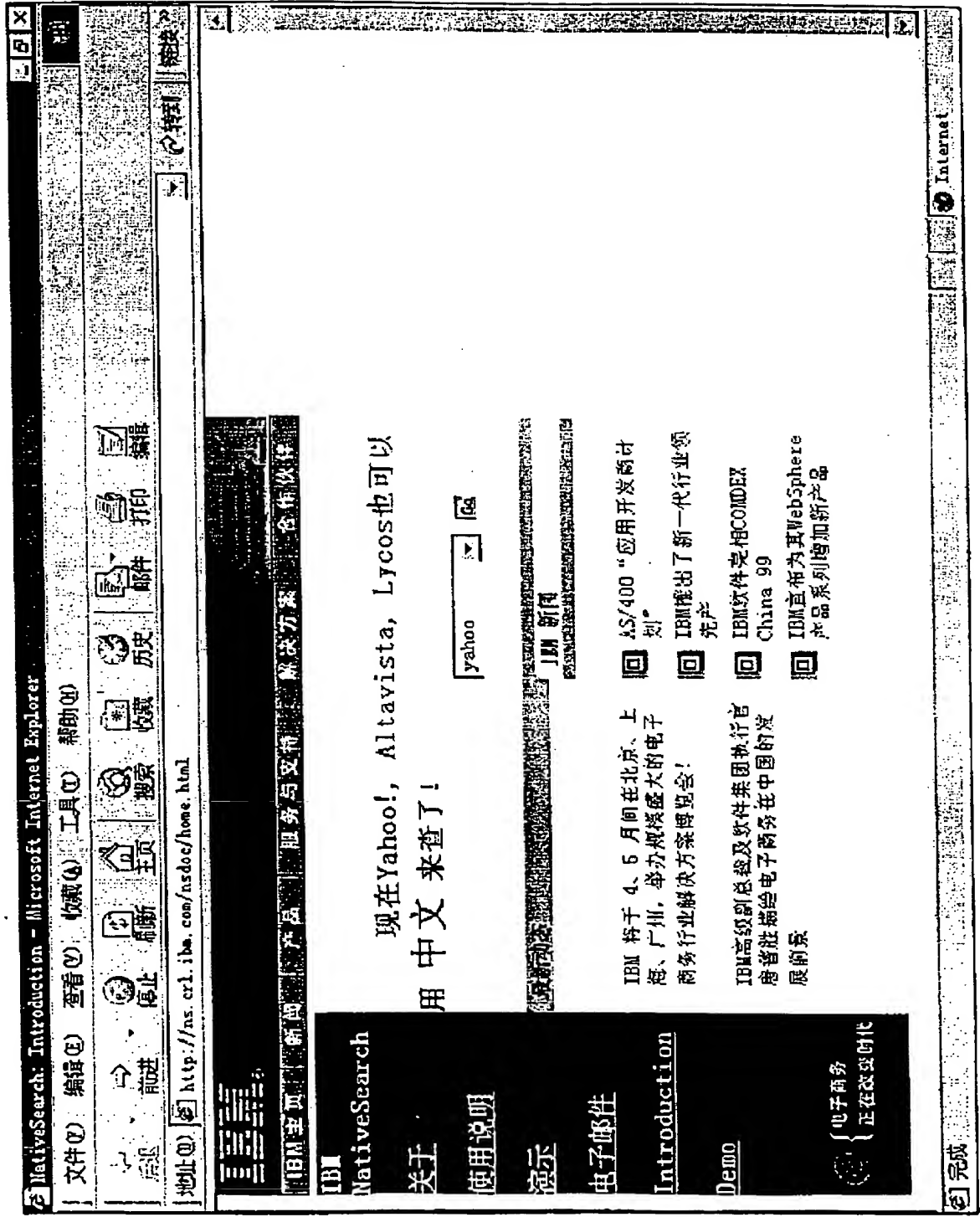


图 7

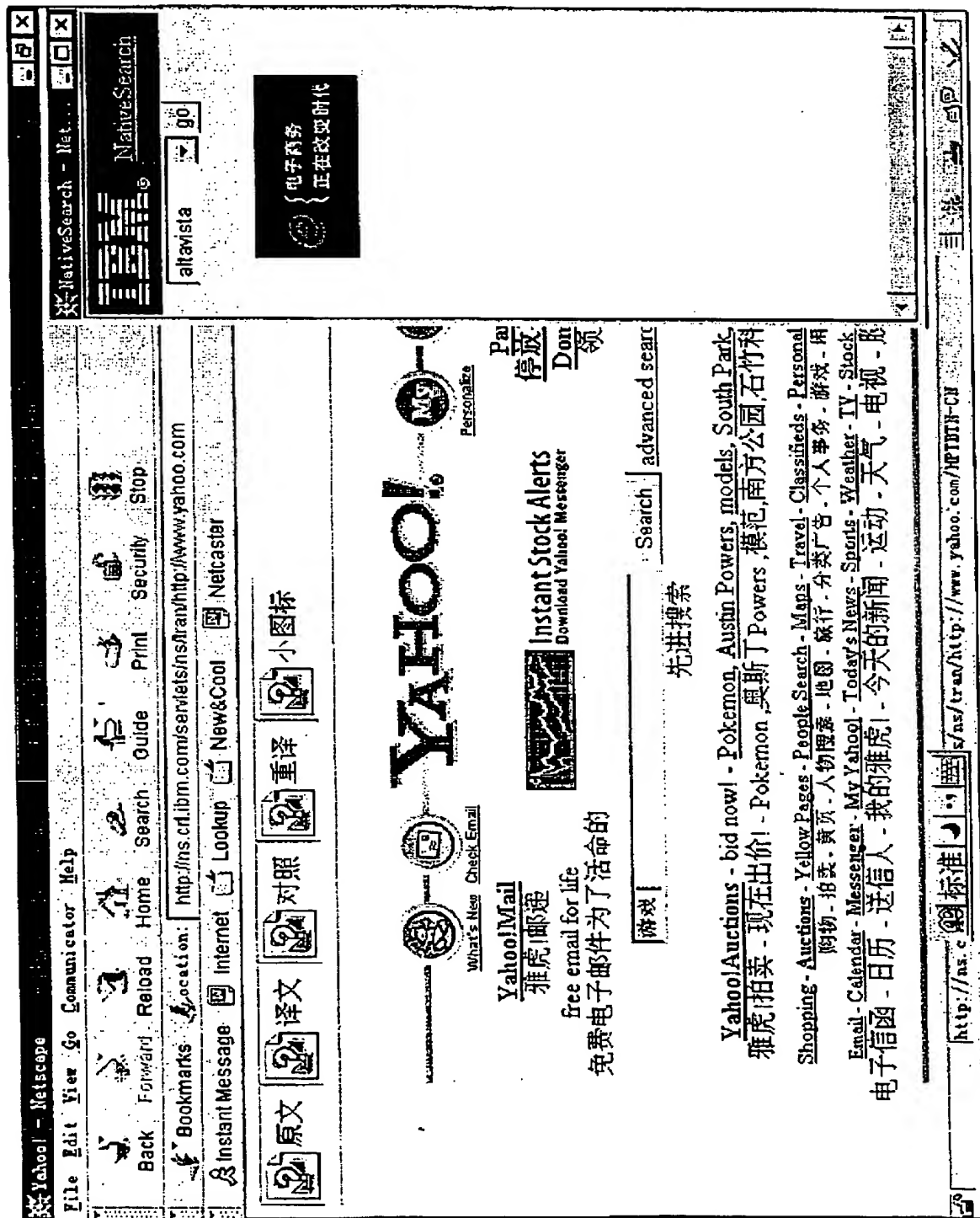


图 8

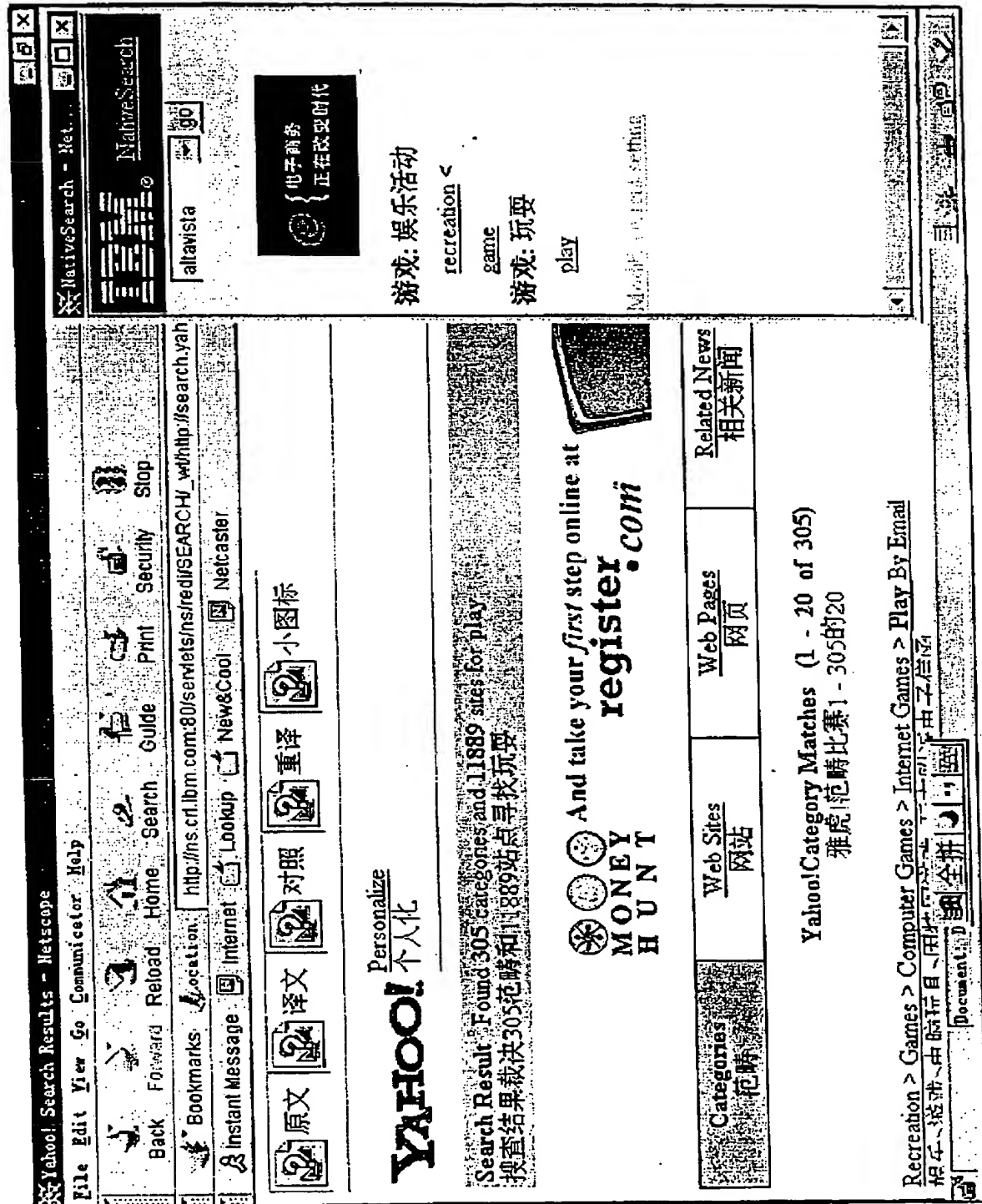


图 9

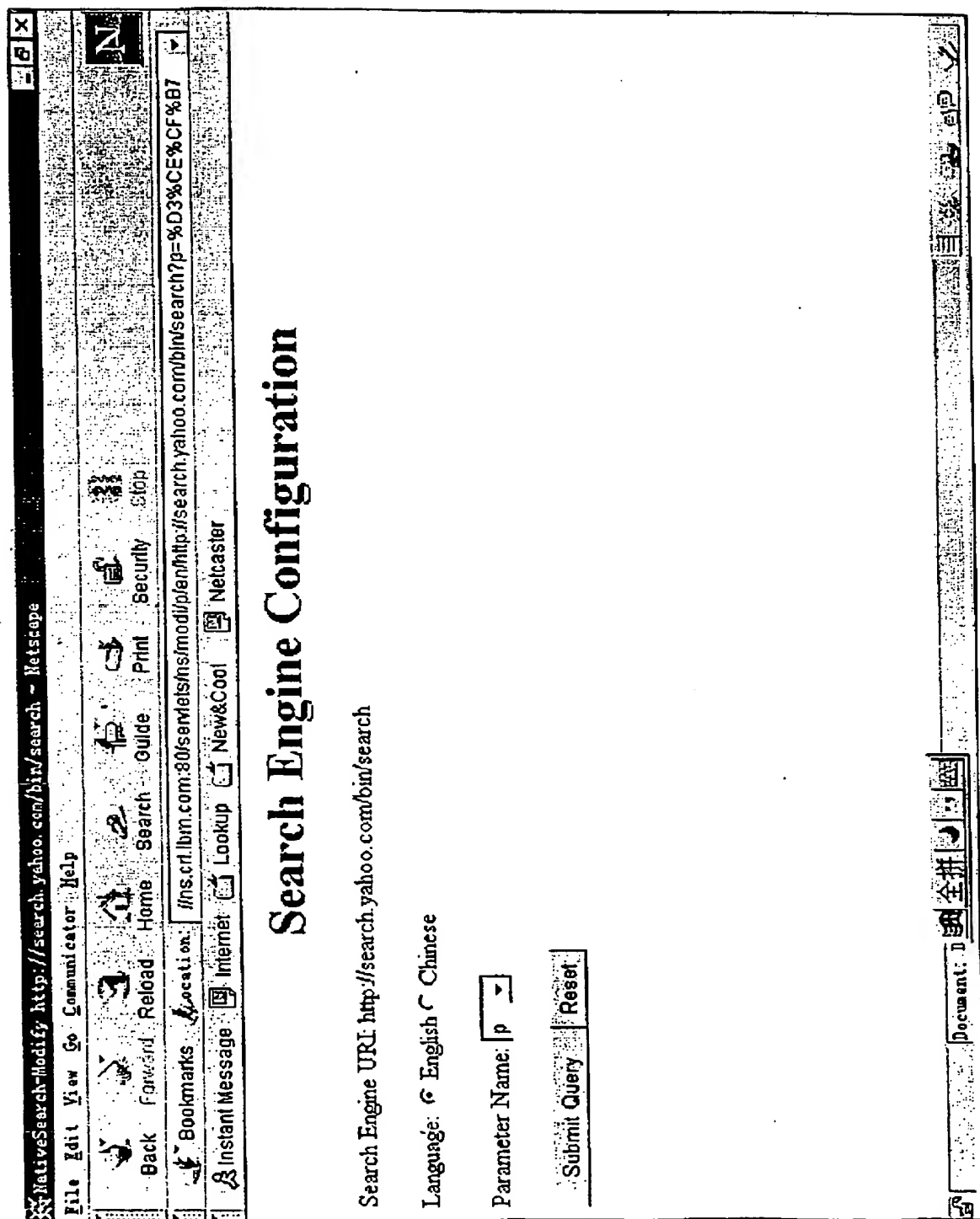


图 10

**This Page is Inserted by IFW Indexing and Scanning
Operations and is not part of the Official Record**

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☐ **BLACK BORDERS**
- ☐ **IMAGE CUT OFF AT TOP, BOTTOM OR SIDES**
- ☐ **FADED TEXT OR DRAWING**
- ☐ **BLURRED OR ILLEGIBLE TEXT OR DRAWING**
- ☐ **SKEWED/SLANTED IMAGES**
- ☐ **COLOR OR BLACK AND WHITE PHOTOGRAPHS**
- ☐ **GRAY SCALE DOCUMENTS**
- ☐ **LINES OR MARKS ON ORIGINAL DOCUMENT**
- ☒ **REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY**
- ☐ **OTHER:** _____

IMAGES ARE BEST AVAILABLE COPY.

As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.